A Framework for the Biomedical Informatics Curriculum

Stephen B. Johnson, Ph.D.
Department of Biomedical Informatics, Columbia University, New York, NY

ABSTRACT

The problem of developing a curriculum for biomedical informatics is highly dependent on how we choose to define and practice the field. Numerous authors have questioned how to position biomedical informatics along the continuum of formal, empirical and engineering disciplines. A concern with current educational programs in biomedical informatics is that students finish without a clear understanding of the relation between theory and practice, or worse, with the impression that the field does not possess any theoretical basis. In this paper, we propose that biomedical informatics curricula explicitly address skills and competencies at three levels: formal, empirical, and applied. We posit that that knowledge of formalization is necessary to build testable empirical models, and that model-driven approaches are necessary for deploying information systems that can be evaluated in a meaningful way. A curricular framework is proposed that identifies a set of methods, techniques and theories that have broad applicability within the domain of biomedicine, and which can span a wide range of application areas: bioinformatics, imaging informatics, clinical informatics and public health informatics. A stronger linkage between theory and practice will result in students who are empowered to create effective and lasting solutions to biomedical problems.

INTRODUCTION

The problem of developing a curriculum for biomedical informatics is highly dependent on how we choose to define and practice the field. Insight can be gained from the approaches taken in established academic disciplines. It is traditional to divide the sciences into the *basic*, such as physics, chemistry, and geology, and the *applied*, which include the various subfields of engineering and many aspects of medicine [1]. This definition implies a continuum, ranging from the pure abstract world of mathematics, to empirical studies, to applications that utilize this knowledge to solve practical problems.

Biomedical Informatics has been positioned at various points along this continuum: as an engineering discipline concerned with developing and evaluating systems [2]; as a modeling discipline concerned with developing formal representation and problem solving methods [3]; as a local science that attempts to explain aspects of a domain in order to design and implement artifacts [4]; and as a broad field that ranges from model formulation, to system development and installation, to the study of their effects [5].

Computer science has conducted a similar self-examination, and produced a similar range of views, with some researchers stating that the field is "not a science, but a synthetic, an engineering discipline" [6], while others taking the position that, for builders of software systems, "We are in fact experimental scientists" [7]. The computer science community has proposed means to harmonize these positions by acknowledging that the field spans the entire continuum [8]. In biomedical informatics, Maojo recommends that we adopt this model, dividing the continuum into three areas: theory (mathematical constructs), abstraction (empirical validation of models) and design (implementation and assessment of systems) [9]. This definition resonates with the major components of those given above. The model also serves as an excellent framework with which to assess current curricula for biomedical informatics, and to direct innovations.

In this paper, we propose that biomedical informatics curricula explicitly address the entire scientific spectrum, from theory to application. The key point is that skills in formalization are necessary to build testable empirical models, and an understanding of experimentation is necessary for deploying information systems that can be evaluated in a meaningful way. In building this educational framework, we seek to identify "a set of methods, techniques and theories that have broad applicability" within the domain of biomedicine [10]. The resulting set of general informatics skills and competencies are intended to span a wide range of application areas: bioinformatics, imaging informatics, clinical informatics and public health informatics.

METHODS

The curricular structure was developed by synthesizing several different kinds of materials, which are discussed in the following sections. As discussed above, published definitions of the field of biomedical informatics are a major source of insight. Current biomedical informatics programs demonstrate various ways of implementing these definitions (this was limited to American programs that offer doctoral degrees). The International Medical Informatics Association (IMIA) has made recommendations on education in health and medical informatics [11], which reflect many of the characteristics seen in current programs. It is also extremely helpful to make comparisons with closely related fields, such as computer science and public health.

Informatics Curricula

Many educational programs in biomedical informatics evolved from a single survey course providing an overview of the field, with selected courses taught through other departments, typically computer science. The survey course introduces students to many interesting problems in biomedicine (usually in patient care), and discusses solutions using computational approaches. Some theory is presented, such as decision making or ontology. There is usually little discussion of experimental methods and significant empirical discoveries of the field.

In subsequent informatics courses, a student may learn a mixture of practice and theory. For example, a course in decision support systems may describe practical applications in patient care and also cover elements of decision theory. A course on controlled vocabulary may describe current coding schemes for billing and diagnosis and also cover formal methods for ontology and description logic. More recently, programs may include a course on evaluation, which typically focuses on assessing the impact of clinical systems.

These courses still leave a significant body of knowledge to impart. Most departments choose to supplement their courses with offerings in other departments. While this avoids duplicate efforts across the university, there is usually a significant difference in focus between the two approaches. For example, a graduate course in databases as taught by computer science may concentrate on abstractions such as relational algebra and algorithms for memory management rather than the complex design and management issues involved in developing large databases in real-world settings. There are also major obstacles of prerequisites and continuity – courses in other departments are designed as part of a larger whole, subject to a particular educational philosophy. Informatics programs have difficulty finding external courses at the right level of detail, and their students may even be prohibited from taking courses in other departments or schools.

At the present time, biomedical informatics programs at most institutions are collections of courses, rather than curricula based on formal educational principles. There is no explicit recognition of the formal, empirical and applied competencies that must be attained. Many students complete such programs without a clear understanding of the relation between theory and practice, or worse, with the impression that the field does not possess any theoretical basis. As a result, students have a hard time perceiving biomedical informatics as a science.

IMIA Recommendations

IMIA’s recommendations on education in health and medical informatics provide a framework to support the development of courses, course tracks and programs. The guidelines address the educational needs of health professionals as well as health and medical information “specialists”, who are the focus of this paper.

The recommendations distinguish three “domain areas”, which may be characterized as follows: familiarity with biomedicine (medicine, health, biosciences); formal knowledge and skills (computer science, mathematics, biometry); and applied methods and techniques for information processing in medicine and health care.

The applied methods include advanced knowledge and skills in systems architecture, data representation and coding, medical records, decision-making and data analysis. There is greater emphasis on clinical applications, but advanced knowledge of bioinformatics, bioimaging, and public health is recommended for certain professionals, which corresponds to the notion of specialization tracks.

The formal knowledge and skills include statistics and logic, as well as some computer science theories. The area described as “biometry” includes introductory knowledge and skills of study design and evaluation methods. There is no explicit guidance on the use of formal knowledge to develop empirical models, or on the application of such models to implement practical systems.

(a)				(b) Prerequisites
Applied				biomedical science (molecular biology, anatomy, physiology, or public health)
Bioinformatics (molecules/cells)	Bioimaging (tissues/organs)	Clinical Informatics (patients)	Public Health Informatics (populations)	
Empirical Empirical methods and theories pertaining to cognitive, behavioral and organizational aspects of information systems				statistics, elementary experimental methods
Formal Mathematical and technical methods and theories				introductory computer science (programming language, data structures, algorithms)

Table 1. Formal, empirical and applied competencies, and associated prerequisites

Related Disciplines

Biomedical informatics is still a young field, and can benefit from comparison to more mature disciplines. Computer science and public health are also relative newcomers to science, but have passed through the growing pains of defining their identities and specifying core competencies.

Biomedical informatics clearly depends on computer science for certain formal aspects as well as more practical engineering approaches. Unlike informatics, formalism in computer science can be an end in itself. This is seen in computability and complexity (the two major pillars of computer science), where mathematical proof serves as the primary method of validation. The two fields share the goal of building efficient and effective systems, but computer science is less interested in empirical methods. Experimental design is not a component of most computer science curricula, and few programs require statistics. Elective courses such as user interface design provide some exposure to cognitive science. Similarly, software engineering can increase awareness of people and organizational issues. However, the focus tends to be on the technology rather than empirical approaches.

Public health provides a complementary model. Its formal methods are grounded in statistics and measurement theory. More importantly, public health also has a rich repertoire of established empirical models. Like informatics, public health is an applied science that draws on existing theories from many disciplines, and applies them to health problems. The two fields share the same challenges of developing curricula for a diverse body of students who may ultimately specialize in a very wide variety of applications. However, unlike informatics, public health has successfully articulated a set of core competencies that constitute a unique scientific discipline.

RESULTS

The framework of Maojo et al. [9] offers a comprehensive approach to informatics that spans from theory to practice. The model consists of three levels, which we adapt as follows:

1. **Formal** – mathematical and technical methods and theories.

2. **Empirical** –methods and theories pertaining to cognitive, behavioral and organizational aspects of information systems, built on the formalisms and technologies of level (1).

3. **Applied** – use of the models and theories of level (2) to solve problems in biology, physiology, patient care, and health.

While the ultimate goal of biomedical informatics is the development and delivery of applications, it is crucial that this work be conducted in a scientific manner, by implementing and deploying established empirical models. In additional, students must understand how to develop such models by drawing on the formalisms that underlie them.

The framework can be used to contrast biomedical informatics with its related fields. Computer science is largely focused on formal and technical issues (level 1), with some forays into experimental models

	Data	Knowledge	Systems
Empirical (cognitive, behavioral, organizational)	data modeling, visualization, data standards, utilization modeling (ownership and authorization)	knowledge acquisition, decision analysis, computer-aided instruction	user interfaces, project management, needs assessment, impact assessment
Formal (mathematical, technical)	ontology, databases, signal processing, image processing, sequence analysis/parsing	data mining, knowledge bases, rule bases, information retrieval	systems architecture, system integration, software engineering

Table 2. Framework of informatics core competencies, with examples

(level 2), e.g., user interfaces. This can be contrasted with what might be termed "general" informatics, which does not focus on a particular domain of application, such as medicine or law. This discipline shares many of its formal methods with computer science (level 1), but has a strong emphasis on cognitive and behavioral models (level 2). Applied sciences (e.g. public health) tend to have somewhat less focus on formalism (level 1), establish a core of experimental methods and models (level 2), and apply these to solving problems in a specific domain (level 3), such as reducing health risks and exposures. Biomedical informatics places similar emphasis on applications.

This framework embodies the philosophy that there are general formal and empirical methods that apply across all biomedical domains: bioinformatics (molecule/cell), bioimaging (tissue/organ), clinical informatics (patients), and public health informatics (populations). For example, database design is a fundamental and general competency, and can be applied to storing biological sequences, images of organs, patient encounters, or epidemiologic surveys. The dependencies of the framework are depicted in Table 1 (a).

The framework is also helpful for identifying the prerequisites that enable acquisition of the proposed informatics skills and competencies. Table 1 (b) shows the associated prerequisites for each level of the model. To begin mastery of formal skills and competencies, students require exposure to introductory computer science (programming language, data structures, algorithms). Empirical training in informatics requires knowledge of statistics, at minimum, as well as familiarity with the scientific method. To effectively apply this knowledge in a particular domain, students need some introduction to an area of biomedical science (e.g., molecular biology, anatomy, physiology, public health). These prerequisites can be defined as required for entrance to the program, or can be imparted within the program at an early stage.

The formal and empirical competencies form a core, which requires further structure. For this, it is useful to consider the varieties of artifacts that serve as objects of informatics study. These artifacts fall along a continuum of increasing complexity, in which three divisions can be made: data, knowledge, and systems. This provides a useful structure for organizing core competencies, in which mathematical and technical issues are distinguished from cognitive, behavioral and organizational aspects (Table 2).

DISCUSSION

The proposed curriculum can be implemented as specific course sequences in several ways. A typical program for predoctoral study requires three years. Dividing Table 1 horizontally, the program would progress from theory to application. The first year would focus on formal and technical methods, the second on empirical techniques and theories, and the third on an application area in biomedicine (with some mixture and overlap across these levels). Dividing Tables 1 and 2 vertically, the program would progress from data to systems, integrating theory and practice in each year. The first year would focus on formal, empirical and applied aspects of data, the second on biomedical knowledge, and the third on complete information systems.

Students entering biomedical informatics come from a very wide range of backgrounds, with greatly varying degrees of preparation in each of the prerequisite areas identified in Table 1 (b). For example, students with a medical background will easily meet the prerequisites for applied competencies, while computer science students will satisfy many of the formal requirements. Leveling of these differences can be achieved by addressing deficiencies in the first year of study, or by requiring students to supplement their education prior to entry. The ultimate goal of the framework is to minimize background differences by the second or third year by immersing students in a coherent educational structure.

At the present time, most informatics programs place greater emphasis on formal and technical methods rather than empirical techniques and theories. For example, some programs have students take a course in databases (taught either by the department or by computer science). However, it is extremely rare that students receive training in data modeling, which involves experimental techniques for eliciting data properties and needs from users, model development and validation. Similarly, there are many courses on decision analysis or decision support, but few on knowledge acquisition, based on experimental methods and models drawn from cognitive science.

A key innovation of the proposed framework is recognition of biomedical informatics as a science (rather than just system building), which some have described as a "modeling discipline" [12]. This view incorporates the need for general, formal models such as ontology [3]. Students must also understand how to develop new models using formal techniques in their chosen biomedical domains, such as clinical guidelines [13]. Finally, it is essential that students acquire a repertoire of established empirical models, and learn how to apply these to drive system implementation and assess impact [5].

CONCLUSION

The primary motivation of the proposed curriculum is to provide an underlying structure for biomedical informatics programs, to move beyond collections of unrelated courses, and to reduce redundancies. The methodologist Donald T. Campbell observed, "Interdisciplinary programs have been misled by goals of breadth and multidisciplinary training" [14], and held that a science composed of multiple disciplines must form a continuous texture of narrow specialties that overlap with one another. Overlap is unavoidable in complex fields such as public health and biomedical informatics. However, the constituent disciplines can be defined to cover the field more efficiently. The proposed curricular framework attempts to distribute areas of study more evenly across the continuum of biomedical scales (molecules to populations), the continuum of complexity (data to systems), and the continuum of application (theory to practice).

This framework explicitly presents biomedical informatics as a distinct discipline with a unique set of formalisms, established empirical models, and applications founded on scientific principles. A stronger linkage between theory and practice will result in students who are empowered to create effective and lasting solutions to biomedical problems.

REFERENCES

1. The Columbia Encyclopedia, Sixth Edition. Columbia University Press, 2002.
2. Shahar Y. Medical informatics: between science and engineering, between academia and industry. Methods of Information in Medicine 2002;41(1):8-11.
3. Musen MA. Medical informatics: searching for underlying components. Methods of Information in Medicine 2002;41(1):12-9.
4. Patel VL, Kaufman DR. Science and practice: a case for medical informatics as a local science of design. Journal of the American Medical Informatics Association 1998;5(6):489-92.
5. Friedman CP. Where's the science in medical informatics? Journal of the American Medical Informatics Association 1995;2(1):65-7.
6. Brooks FP. The computer scientist as toolsmith II. Communications of the ACM 1996;39(3):61-68.
7. Plaice J. Computer science is an experimental science. ACM Computing Surveys 1995;27(1):33.
8. Denning P, Conner D, Gries D, Mulder Tucker A, Young P. Computing as a discipline. Communications of the ACM 1989;32(1):9-23.
9. Maojo V, Martin F, Crespo J, Billhardt H. Theory, abstraction and design in medical informatics. Methods of Information in Medicine 2002;41(1):44-50.
10. Shortliffe EH, Johnson SB. Medical informatics training and research at Columbia University. IMIA Yearbook of Medical Informatics 2002.
11. Anonymous. Recommendations of the International Medical Informatics Association (IMIA) on education in health and medical informatics. Methods of Information in Medicine 2002;39(3):267-77.
12. Hasman A. Haux R. Albert A. A systematic view on medical informatics. Computer Methods & Programs in Biomedicine 1996;51(3):131-9.
13. van der Maas AA. ter Hofstede AH. ten Hoopen AJ. Requirements for medical modeling languages. Journal of the American Medical Informatics Association 2001;8(2):146-62.
14. Campbell DT. Ethnocentrism of disciplines and the fish-scale model of omniscience. In M. Sherif & C. W. Sherif (Eds.), Interdisciplinary relationships in the social sciences. Chicago: Aldine Publishing 1969:328-48.